



# HOMELAND SECURITY AND DEFENSE CENTER

CHILDREN AND FAMILIES  
EDUCATION AND THE ARTS  
ENERGY AND ENVIRONMENT  
HEALTH AND HEALTH CARE  
INFRASTRUCTURE AND  
TRANSPORTATION  
INTERNATIONAL AFFAIRS  
LAW AND BUSINESS  
NATIONAL SECURITY  
POPULATION AND AGING  
PUBLIC SAFETY  
SCIENCE AND TECHNOLOGY  
TERRORISM AND  
HOMELAND SECURITY

The RAND Corporation is a nonprofit institution that helps improve policy and decisionmaking through research and analysis.

This electronic document was made available from [www.rand.org](http://www.rand.org) as a public service of the RAND Corporation.

Skip all front matter: [Jump to Page 1](#) ▼

## Support RAND

[Purchase this document](#)

[Browse Reports & Bookstore](#)

[Make a charitable contribution](#)

## For More Information

Visit RAND at [www.rand.org](http://www.rand.org)

Explore the [RAND Homeland Security and Defense Center](#)

View [document details](#)

## Limited Electronic Distribution Rights

This document and trademark(s) contained herein are protected by law as indicated in a notice appearing later in this work. This electronic representation of RAND intellectual property is provided for non-commercial use only. Unauthorized posting of RAND electronic documents to a non-RAND website is prohibited. RAND electronic documents are protected under copyright law. Permission is required from RAND to reproduce, or reuse in another form, any of our research documents for commercial use. For information on reprint and linking permissions, please see [RAND Permissions](#).

Report Documentation Page				Form Approved OMB No. 0704-0188	
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE <b>2011</b>		2. REPORT TYPE		3. DATES COVERED <b>00-00-2011 to 00-00-2011</b>	
4. TITLE AND SUBTITLE <b>Measuring Illegal Border Crossing Between Ports of Entry: An Assessment of Four Promising Methods</b>				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) <b>Rand Corporation, Homeland Security and Defense Center, 1776 Main Street, PO Box 2138, Santa Monica, CA, 90407-2138</b>				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT <b>Approved for public release; distribution unlimited</b>					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT <b>Same as Report (SAR)</b>	18. NUMBER OF PAGES <b>26</b>	19a. NAME OF RESPONSIBLE PERSON
a. REPORT <b>unclassified</b>	b. ABSTRACT <b>unclassified</b>	c. THIS PAGE <b>unclassified</b>			

This product is part of the RAND Corporation occasional paper series. RAND occasional papers may include an informed perspective on a timely policy issue, a discussion of new research methodologies, essays, a paper presented at a conference, a conference summary, or a summary of work in progress. All RAND occasional papers undergo rigorous peer review to ensure that they meet high standards for research quality and objectivity.



# Measuring Illegal Border Crossing Between Ports of Entry

An Assessment of Four  
Promising Methods

Andrew R. Morral, Henry H. Willis,  
Peter Brownell

Approved for public release; distribution unlimited



The research described in this report was conducted within the Homeland Security and Defense Center, a joint center of RAND Infrastructure, Safety, and Environment and the RAND National Defense Research Institute, a federally funded research and development center sponsored by the Office of the Secretary of Defense, the Joint Staff, the Unified Combatant Commands, the Navy, the Marine Corps, the defense agencies, and the defense Intelligence Community under Contract W74V8H-06-C-0002.

**Library of Congress Cataloging-in-Publication Data** is available for this publication.

ISBN: 978-0-8330-5150-9

The RAND Corporation is a nonprofit institution that helps improve policy and decisionmaking through research and analysis. RAND's publications do not necessarily reflect the opinions of its research clients and sponsors.

**RAND®** is a registered trademark.

© Copyright 2011 RAND Corporation

Permission is given to duplicate this document for personal use only, as long as it is unaltered and complete. Copies may not be duplicated for commercial purposes. Unauthorized posting of RAND documents to a non-RAND website is prohibited. RAND documents are protected under copyright law. For information on reprint and linking permissions, please visit the RAND permissions page (<http://www.rand.org/publications/permissions.html>).

Published 2011 by the RAND Corporation  
1776 Main Street, P.O. Box 2138, Santa Monica, CA 90407-2138  
1200 South Hayes Street, Arlington, VA 22202-5050  
4570 Fifth Avenue, Suite 600, Pittsburgh, PA 15213-2665  
RAND URL: <http://www.rand.org>  
To order RAND documents or to obtain additional information, contact  
Distribution Services: Telephone: (310) 451-7002;  
Fax: (310) 451-6915; Email: [order@rand.org](mailto:order@rand.org)

## Preface

---

The U.S. Department of Homeland Security (DHS) is responsible for securing the land, air, and maritime borders of the United States. Strategic planning is necessary if DHS is to meet its responsibilities effectively and efficiently. As part of that planning, DHS leadership must define concrete and sensible objectives and measures of success. These can be used to assess results along the way, to guide allocation of resources, and to inform programming and budgeting for future capabilities and functions.

To support these efforts, the RAND Corporation was asked by the DHS Office of Program Analysis and Evaluation for research and recommendations about strategic-level measures for assessing the effectiveness of border security efforts. The resulting report, *Measuring the Effectiveness of Border Security Between Ports-of-Entry* (Willis, Predd, Davis, and Brown, 2010), highlighted the importance of developing better estimates of the flow of illegal immigrants across the border. We now elaborate in this report on several innovative methods for deriving estimates of the flow of illegal immigration that have not yet fully been explored.

This paper is part of the series *New Ideas in Homeland Security*, a set of RAND Corporation research papers on fundamental questions of homeland security in the United States. The series began during the transition between presidential administrations in 2008–2009, a period when approaches to homeland security were being reassessed. Each paper explores different approaches to ongoing homeland security policy problems. In doing so, they frame the kinds of questions that must be considered if policies shaping homeland security are to be effective. Questions in this area include the following: Are current practices the best and most economical ones for keeping the United States safe? Are better means available for evaluating what may work or not and why?

This series is designed to focus on a small set of policy areas, produce essays exploring different approaches to ongoing policy problems, and frame questions that need to be answered if homeland security policy is to be improved. The resulting discussions should be of interest to homeland security policymakers at the federal, state, and local levels and to members of the public interested in homeland security and counterterrorism.

Earlier papers in this series include

- Brian A. Jackson, *Marrying Prevention and Resiliency: Balancing Approaches to an Uncertain Terrorist Threat*, Santa Monica, Calif.: RAND Corporation, OP-236-RC, 2008.
- Brian A. Jackson, *The Problem of Measuring Emergency Preparedness: The Need for Assessing “Response Reliability” as Part of Homeland Security Planning*, Santa Monica, Calif.: RAND Corporation, OP-234-RC, 2009.

- C. Richard Neu, *Is It Time to Rethink U.S. Entry and Exit Processes?* Santa Monica, Calif.: RAND Corporation, OP-235-RC, 2009.
- Brian A. Jackson and David R. Frelinger, *Emerging Threats and Security Planning: How Should We Decide What Hypothetical Threats to Worry About?* Santa Monica, Calif.: RAND Corporation, OP-256-RC, 2009.
- Brian A. Jackson and David R. Frelinger, *Understanding Why Terrorist Operations Succeed or Fail*, Santa Monica, Calif.: RAND Corporation, OP-257-RC, 2009.
- Andrew R. Morral and Brian A. Jackson, *Understanding the Role of Deterrence in Counterterrorism Security*, Santa Monica, Calif.: RAND Corporation, OP-281-RC, 2009.

## **The RAND Homeland Security and Defense Center**

This research was conducted within the Homeland Security and Defense Center, a joint center of RAND Infrastructure, Safety, and Environment and the RAND National Defense Research Institute, a federally funded research and development center sponsored by the Office of the Secretary of Defense, the Joint Staff, the Unified Combatant Commands, the Navy, the Marine Corps, the defense agencies, and the defense Intelligence Community.

For more information on the Homeland Security and Defense Center, see <http://www.rand.org/multi/homeland-security-and-defense.html> or contact the director (contact information is provided on the web page).

# Contents

---

<b>Preface</b> .....	iii
<b>Summary</b> .....	vii
<b>Abbreviations</b> .....	ix
CHAPTER ONE	
<b>Why Estimate Flow?</b> .....	1
CHAPTER TWO	
<b>Four Promising Methods</b> .....	3
Improved Capture-Recapture Methods .....	3
Stratified Sampling of Border Crossings .....	5
Surveys and Respondent-Driven Sampling .....	7
Synthetic and Proxy Measures .....	9
CHAPTER THREE	
<b>Discussion</b> .....	11
<b>References</b> .....	13





## Summary<sup>1</sup>

---

DHS is responsible for controlling the flow of goods and people across the U.S. border, a difficult task raising challenging resource management questions about how best to minimize illicit flows across the border while facilitating legitimate ones. Ideally, DHS could evaluate alternative investment strategies in terms of objective measures of the effectiveness with which border control systems prevent illicit flows or perform other essential functions. Instead, commonly reported border control measures, such as numbers of illegal migrants apprehended or miles of border under effective control, bear only an indirect and uncertain relationship to the border control mission, making them unreliable management tools.

Fundamental to the question of border control effectiveness is the proportion of illicit border crossings that are prevented either through deterrence or apprehension. Estimating these proportions requires knowing the total flow of illicit goods or border crossings, but compelling methods for producing such estimates do not yet exist. The objective of this short paper is to describe four innovative approaches to estimating the total flow of illicit border crossings between ports of entry. Each is sufficiently promising to warrant further attention for purposes of supporting reliable, valid, and timely measures of illicit cross-border flow. These are

- improved capture-recapture methods that account for attrition in the pool of migrants attempting to cross after an apprehension and of the number and length of successful crossings migrants make between apprehensions
- estimating total crossings from a stratified sampling of crossings at border segments selected with a known probability
- respondent-driven sampling surveys of migrant communities in the United States
- synthetic modeling either of the stock and flow of the migrant community or of the black market economy for coyote services.

Successfully implementing each of these approaches will require methodological development and analysis to identify barriers or constraints to using the approach, the cost of data collection, and the amount of error that can be expected in the resulting estimates.

---

<sup>1</sup> We would like to thank Joel Predd and Greg Ridgeway, who reviewed and made substantive contributions to this document.



## Abbreviations

---

CBP	U.S. Customs and Border Protection
DHS	U.S. Department of Homeland Security
EMIF	<i>Encuesta sobre Migración en la Frontera Norte de México</i>
GAO	U.S. Government Accountability Office
ONDCP	Office of National Drug Control Policy



## Why Estimate Flow?

---

Movement of goods and people across the borders is critical to the U.S. economy and to the quality of life in communities across the country. Each year over 21 million containers bring products and components to stores and factories (CBP, 2010), and more than 47 million people visit the United States from abroad for tourism and business (ITA, undated). However, in addition to this legitimate travel, goods and people flow across the border illicitly, contributing to significant national problems, such as the availability of illegal drugs, threats of terrorism, and growth in the population of what may be 11.1 million undocumented migrants, many of whom entered the country between ports of entry (DHS, 2010; Passel and Cohn, 2010).

The U.S. Department of Homeland Security (DHS) is principally responsible for controlling the flow of goods and people across the U.S. border, a difficult task raising challenging questions about how best to use technology, intelligence, and manpower to minimize illicit flows across the border while facilitating legitimate ones. Ideally, DHS leadership could evaluate alternative investment strategies in terms of objective measures of the effectiveness with which border control systems prevent illicit flows or perform other essential functions. Instead, commonly reported border control measures, such as numbers of illegal migrants apprehended or miles of border under effective control, bear only an indirect and uncertain relationship to the border control mission, making them unreliable management tools. For instance, increases in border control effectiveness could well increase the number of apprehensions, without such increases signaling greater flow of illegal crossers. Interestingly, U.S. Customs and Border Protection (CBP, 2009) has reported annual *decreases* in apprehensions as evidence of effectiveness, presumably on the theory that decreases reflect fewer crossing attempts, rather than diminished apprehension success rates. Indeed, the U.S. Government Accountability Office (GAO, 2009) reports that CBP explained *increases* in apprehensions made at checkpoints in some border sectors to improved CBP operations and *decreases* in apprehensions in other sectors to the deterrent effects of improved CBP technologies and increased staffing. Clearly, a measure that reflects successful performance whether it rises or falls has limited value as a management tool.

Similarly, observed deaths of those attempting illegal crossings would be a valid proxy for total flow if the probability of death were constant for crossers over time and independent of border control effectiveness. Neither condition is likely true. The per crossing risk of death is a function of the place where crossings occur and the transportation means used, both of which vary over time and may be driven by border control efforts. As easier means of crossing are effectively interdicted, crossers may have to accept higher fatality risks. As such, increases in deaths could reflect either a stable risk of death but increased flow or a stable flow but increased risk of death. Moreover, risk of death is also determined by other factors outside of the control

of DHS, such as extreme temperatures and flash flooding (Eschbach et al., 1999). Without information distinguishing these and similar possibilities, deaths, like apprehensions, cannot reasonably be used to assess the effectiveness of border control.

Similarly, comparable measures, such as apprehensions per patrol agent work year, a measure suggested by GAO (GAO, 2005), suffers the same limitations. The problem with all such measures, as well as with such currently favored measures as miles of border under effective control, is correctly summarized by GAO (2009, p. 29), which says that such measures have been shown to “bear little relationship to effectiveness because they do not compare these numbers to the amount of illegal activity that passes through undetected.”

In work sponsored by DHS, RAND proposed that border security performance requires metrics that address three high-level functions: the proportion of illegal border crossings that are interdicted, the extent to which those who would otherwise cross the border are deterred, and the extent to which border security systems make use of and generate valuable intelligence (Willis et al., 2010). Knowing what to measure is different than knowing how to measure it, and each of the RAND metrics poses difficult methodological challenges for generating measurements that are reliable, valid, sensitive, and sufficiently timely to be used for management decisions.

Good estimates of the total flow of illicit traffic between ports of entry lay at the heart of each of RAND’s metrics but most straightforwardly so for interdiction rates, where total flow is required for the denominator. Because reliable and timely estimates of illicit flow do not currently exist, the objectives of this short concept paper are to propose four innovative approaches to measuring flow and to argue that they are sufficiently promising for use in performance measurement to merit additional research and development.

We limit discussion here to the problem of estimating flow of migrants into the United States between ports of entry because fairly good methods exist to measure illicit flows at ports of entry. For instance, CBP calculates an expected flow of illegal immigrants through ports of entry based on a thorough inspection of a random selection of vehicles, pedestrians, and air travelers entering the country. Data from this Compliance Examination (COMPEX) program can be used to reliably estimate the proportion of illegal flow that goes undetected at ports of entry. In fiscal year 2006, for instance, CBP estimated that several thousand inadmissible aliens successfully entered the country through land or air ports of entry (GAO, 2007).

The remainder of this paper is organized around four methods for estimating flow between ports of entry: measures of flow collected at the border via improved capture-recapture methods and stratified sampling of border crossings, measures collected on migrant populations, and synthetic and proxy measures of illegal border crossings.

## Four Promising Methods

---

Border control agencies and other existing or new data collection systems offer considerable promise for evaluating border control performance and the flow of migrants crossing into the United States between ports of entry. At present, many of the existing data streams are inadequate for management purposes, but changes to collection and analysis strategies could yield valuable information about the flow of illegal migrants. In this section we describe four methods for modifying how data is collected and analyzed that appear sufficiently promising to merit further research and development.

### Improved Capture-Recapture Methods

Bird and rare animal populations are sometimes estimated using a capture-recapture method: A random sample of individuals in a population are captured, tagged, and released. Then a second, independent random sample is captured, from which the proportion found to have previously been tagged provides an estimate of the probability of being sampled,  $p$ , from which the size of the overall population can be calculated as

$$N = (\text{number of tagged individuals})/p.$$

Similar approaches have been proposed for estimating the volume of illegal crossers. For instance, Espenshade (1995) proposes a “repeated trials” approach that considers the population of those arrested trying to cross the border, and specifically the number of times arrestees have been recaptured after prior attempts to cross. Espenshade shows algebraically that if border crossers in any given month continue trying to cross the border until they succeed, the monthly probability of arrest for each crossing attempt,  $p$ , equals the ratio of those arrested who were previously denied entry (i.e., repeat crossers,  $R$ , on their second or later attempts to cross) to all arrestees,  $A$ . That is,  $p = R/A$ . Knowing  $p$  and  $A$ , it is therefore possible to calculate the total monthly flow of illegal crossings,  $F$ , as  $F = A(1-p)/p$ .

Key shortcomings of this approach are that it assumes that everyone who tries to cross the border will continue to attempt crossing until they succeed and that those with prior arrests have not succeeded in entering the United States between arrests (i.e., they have not successfully crossed, returned to Mexico or another country, and are now on a return trip to the United States). Whereas many migrants do continue to try to cross until they succeed (Donato, Durand, and Massey, 1992), some portion of detained crossers will be deterred from further attempts because of the risks posed by hazardous transport routes, the discomfort of border



patrol detention and processing, or other costs of each crossing attempt. Similarly, large numbers of illegal seasonal migrants, those working as coyotes, or those cycling between the United States and another country for other reasons could undermine a key assumption required for the proportion of repeat arrests to all arrests to represent the probability of arrest. Without information about the deterrent effects of arrest and deportation and the number of successful crossings between arrests, the probability of arrest and the total flow cannot be calculated.

Chang et al. (2006) develop Espenshade's capture-recapture model to explicitly account for the missing deterrence (or "discouragement"),  $D$ , and "return and reentry,"  $R$ , factors. However, without information about the magnitude of these variables, Chang and colleagues instead conduct a sensitivity analysis, in effect answering the question, "How great might flow be if  $D$  and  $R$  range between 0 and 20 percent of the total population?" However, whether 20 percent is a reasonable upper bound on these parameters is unknown. Similarly, for use as a performance measure, we would want to compare estimates of flow over time or across border sectors, requiring the additional unverified assumption that  $D$  and  $R$  do not vary substantially over time or geographic border segments.

Clearly, the model developed by Chang et al. (2006) would be improved with empirical estimates of  $D$  and  $R$  (deterrence and return and reentry), and data sources exist that might provide this information. For instance, the Mexican Ministry of Labor conducts a probability sample survey of migrants in northern Mexican border communities called *Encuesta sobre Migración en la Frontera Norte de México* (EMIF), which is collected at frequent intervals. One group specifically targeted in this survey is migrants who were recently apprehended and returned to Mexico by U.S. Border Patrol authorities (Rendall et al., 2009). Respondents are queried about their migration history, their most recent migration effort, the number of crossing attempts and apprehensions they experienced, and their future migration intentions. Data from this survey might be useful for estimating attrition from the pool of migrants seeking entry after apprehension by border control and possibly the number of successful entries between apprehensions at the border.

In addition, comparison of demographic and other characteristics of those interdicted and returned by border patrol and those identified as recent returnees in the EMIF could be used to clarify whether the risk of apprehension is constant across diverse groups of migrants—a key assumption for the capture-recapture methods—or whether there are important subgroup differences in the risk of apprehension that suggest the need for separate parameter estimates—e.g., between Mexican migrants and non-Mexican migrants or between those working with a coyote and those crossing on their own.

Finally, better estimates of  $D$  might be calculated from existing apprehension data. The "repeated trials" method predicts that a constant proportion of any cohort will be apprehended in their next crossing attempt. In other words, the probability of being apprehended does not depend on the number of previous apprehensions. However, in practice this may not be true if people are discouraged from attempting multiple times or are able to learn from being apprehended and improve crossing tactics. Thus, suppose that on the basis of those with one and two prior apprehensions we estimate the probability of apprehension to be  $1/3$ , but we observe that for every 1,000 migrants with two apprehensions, only 200 are apprehended a third time. This would suggest that the number of people attempting a third crossing is substantially lower than 1,000 and would in fact be  $200/0.333 = 600$ , suggesting that  $D$  for a third crossing is  $(1,000 - 600)/1000 = 0.40$ . A flaw here is that it implicitly treats  $D$  as zero after the first apprehension.

However, estimates of  $D$  from higher numbers of apprehensions could be used recursively to estimate attrition after the first arrest.

The EMIF provides data that could provide an independent estimate of changes in  $D$  over time. Specifically, it asks recently deported migrants and those who departed voluntarily about their intentions to attempt an entry in the United States. We would expect changes in the proportion of such individuals claiming no intention of a return to the United States to be comparable to changes over time calculated from apprehension data or other methods.

Whereas these estimates could be affected by unknown values of  $R$ , we suspect that with reasonable estimates of how long migrants stay after crossing successfully, repeat apprehension data could be selected over a sufficiently short time frame to make  $R$  of trivial importance to the estimate of flow. Nevertheless, because migrant learning—and success evading capture—produces the same apparent attrition from the recaptured samples, the estimates of  $D$  described here would have to be treated as upper bounds on the proportion of migrants who are discouraged after any specific number of crossing attempts; some portion of those who appear discouraged have, in fact, become more canny, reducing their probability of apprehension.

## Stratified Sampling of Border Crossings

Stratified sampling is a technique for producing population estimates based on samples drawn from groups sampled with a known probability. Implementing this approach must address two methodological questions: (1) With what probability is a segment of a population to be sampled, and (2) what is the probability of detection within the sampled population?

For example, a household survey of drug use might randomly select homes and within homes randomly select one person to interview about drug use. In this sampling scheme, individuals who live by themselves are sampled with certainty when their home is drawn, whereas individuals living with, for instance, three others are only sampled with a 25 percent probability. Without correcting drug use frequencies in the sample for the probability of each individual's selection into it, the estimates would be biased toward the drug use behavior of those who live alone or with few others. As such, estimates of the total number of drug users are constructed by weighting survey responses by the inverse probability of selection into the sample. Similarly, homes need not be selected with equal probability, as it may be preferable to “oversample” homes in areas of particular interest. Again, unbiased population estimates can be achieved if oversampled results are corrected with proper sample weighting. Thus, instead of having only a single weight to adjust for a respondent's probability of selection within a house, such a survey design would also require weighting responses by the inverse probability of the house being included in the sample.

Border patrol operations could be viewed as a form of stratified sampling of illicit border crossings, but one that is not yet adequate to produce unbiased estimates of the flow of such crossings. Instead of sampling houses, the border patrol samples regions of the border because it lacks the resources to monitor all points on the border all the time.

Currently the sampling of regions is not probabilistic but deterministic, so statistical inferences about the total number of crossers cannot be drawn from those observed in the sampled regions. However, border patrol resources could be assigned to regions randomly without loss of efficiency or effectiveness and with the benefit of collecting data that would permit estimation of the total flow of illegal crossers.

For instance, the border could be divided into regions of high, medium, and low risk of illegal entries (or strata). Border patrols could be assigned with certainty or with high probability to high-risk areas, with lower but known probability to medium-risk areas, and with the lowest but nonzero probability to low-risk areas. Knowing the probability of “sampling” each region allows for estimates of the total number of crossers who would have been apprehended,  $A_T$ , had patrol resources been available at every region at all times. For instance, suppose that low-risk regions are assigned patrols 5 percent of the time, medium-risk regions are selected for patrol 25 percent of the time, and high-risk regions have 100 percent coverage. If the numbers of apprehensions made in low-, medium-, and high-risk regions in a given month were 50, 100, and 1,000, respectively, an unbiased estimate of the total number of apprehensions that would have occurred with patrols at every region of the border would be  $A_T = (50/0.05) + (100/0.25) + (1,000/1) = 2,400$ , even though only 1,150 were actually apprehended because of resource constraints.

Such a stratified sampling approach tells one how many illegal border crossings would result in apprehensions if border control resources could be assigned to every segment of the border, but if not all crossers in a patrolled segment are detected, the total flow of illegal migrants will be higher. Thus, if the probability of a crosser being apprehended given the presence of a patrol in a region,  $P(\text{Apprehension} \mid \text{Patrol})$ , is known or can be reasonably estimated, then the estimate of the total number of crossers who would be apprehended had patrols been at all places could be transformed into an estimate of total flow,  $F = A_T / [P(\text{Apprehension} \mid \text{Patrol})]$ . For instance, building on the example above, if patrols have an 80 percent chance of apprehending crossers making an attempt in any region with an assigned patrol, the estimate of 2,400 possible apprehensions would represent 80 percent of the total flow,  $F = 2,400/0.8 = 3,000$ .

In practice, there may be several approaches that could be used to estimate the probability of detection in a sampled area, including periodic “red teaming,” in which migrants or confederates are recruited to attempt crossings in order to establish their likelihoods of detection and apprehension or comparison of estimates resulting from sampling the same area with multiple reconnaissance assets (see, e.g., Thompson, 2004, and GAO, 2008). Independent estimates of  $P(\text{Apprehension} \mid \text{Patrol})$  could be used to produce estimates of discouragement,  $D$ , after the first failed attempt to cross. Specifically, we can generate capture-recapture estimates of the probability of apprehension on a border segment

$$P(\text{Apprehension}) = \frac{R_1}{A_1(1 - D_1)},$$

where  $A_1$  is the total number of individuals apprehended on a first crossing attempt,  $R_1$  is the number reapprehended immediately after the first apprehension, and  $D_1$  is the unknown rate at which crossers are discouraged after a first failed crossing attempt. If we substitute this estimate into the equality

$$P(\text{Apprehension}) = P(\text{Apprehension} \mid \text{Patrol}) * P(\text{Patrol}),$$

in which  $P(\text{Patrol})$  is the probability with which CBP is assigning patrols to be present on a given border segment, we can calculate discouragement as

$$D_1 = 1 - \frac{R_1}{A_1 P(\textit{Patrol}) P(\textit{Apprehension} \mid \textit{Patrol})}.$$

Similarly, independent estimates of  $D_1$  could be used to calculate the probability that patrols in an area will apprehend any crossing attempts there,  $P(\textit{Apprehension} \mid \textit{Patrol})$ .

This general approach to sampling crossers could be extended using available technology where border patrols may not be available. Again on the analogy of detecting bird and rare animal populations, discrete “camera traps” or high-altitude overhead reconnaissance could be randomly distributed across the border to improve estimates of the flow across sections of the border when patrols are elsewhere. Here again, by situating these cameras in accordance with a systematic, probabilistic sampling plan, inferences about the total flow may be possible.

In addition to providing data on the flow of illegal crossers, probabilistic sampling of border regions can improve border control performance by increasing the probability of rapidly detecting new transportation routes and strategies. For instance, Predd and Willis (2010) report simulation results showing that allocating a portion of border patrol resources to sectors randomly, rather than deterministically, can increase and optimize the detection of crossers.

Finally, the sampling approach to measurement has the attractive feature of highlighting the correspondence between resources allocated and foregone apprehensions. That is, the approach lends itself to precise estimates of the tradeoffs between changes in border patrol resources available to allocate to sectors and the proportion of crossers that are apprehended.

## Surveys and Respondent-Driven Sampling

Population surveys, such as the American Community Survey and the U.S. Census, have been used to estimate net migrant populations in the United States (e.g., Passel, Van Hook, and Bean, 2004). From these estimates, the proportion of migrants that is undocumented can be inferred as the remainder after accounting through administrative records for those who have been granted legal citizenship or remain on current visitor visas. As a source of data for measuring the border control performance, however, such estimates are of limited value. Population surveys are expensive; they often occur less than annually, making them at best a poorly responsive performance metric; and they may systematically undercount new immigrants who have not yet found stable housing, a group that may include many of those immigrants who crossed the border without authorization (Rendall et al., 2009). More importantly, it may not be possible to differentiate undocumented immigrants who overstayed visas from those who crossed the border without authorization. Population surveys commonly used to evaluate the size of the unauthorized migrant community do not include questions that could be used to draw this distinction, and it is unlikely, because of known underreporting of illicit and negatively sanctioned behavior, that inclusion of such a question would yield valid responses. As such, little is known about the proportion of the unauthorized migrant population that represents overstayers or unauthorized crossers (Passel, 2005).

A more promising population survey approach is represented by the EMIF, which conducts a random probability sample of migrants in or passing through northern Mexican border towns, interviewing respondents about their migration experiences and intentions. The EMIF specifically targets migrants who are returning from stays in the United States, those who have

been apprehended and returned during attempted entry into the United States, and those who are planning to migrate to the United States in the near term. Here again, however, the data collected through the EMIF has important shortcomings as a source of information on border control performance. In particular, because the EMIF data are collected on the Mexican side of the border, they do not provide direct information on the whole population of migrants who have crossed the border over any specified time frames. Although the data may offer good information about who intends to cross the border in the next 30 days, whether all with this intention actually attempt to cross, when they make such attempts, how they ultimately attempt to cross, and the outcome of their attempts are all unknown. Arguably, these migrants have no incentives not to cross, so all or most expressing the intention to cross soon will attempt to do so. Psychological research in other domains suggests, however, that even among those with strong incentives to behave in accordance with their intentions, large proportions often fail to do so (Prochaska, DiClemente, and Norcross, 1992). As such, more empirical work is required before assumptions can safely be made that all who intend to cross do so.

Ideally, undocumented migrants who have crossed into the United States would provide information on when and how they succeeded in entering the country, but this is a population that can be expected to actively avoid contact with authorities seeking this information, and they might be expected to provide inaccurate information when they are surveyed. Nevertheless, some progress has been made in developing survey techniques that are appropriate for efficiently estimating population parameters for other such “hidden” populations engaged in negatively sanctioned behaviors, such as substance abusers or commercial sex workers. One such technique that merits consideration for estimating features of the undocumented migrant population is respondent-driven sampling (Salganik and Heckathorn, 2004).

Respondent-driven sampling begins with a nonrandom sample of individuals from the population of interest, interviewing them about their characteristics of interest (e.g., when and how they arrived in the United States) and then asking them to distribute invitations to participate in the survey to their friends. The friends, in turn, are interviewed and asked to invite their friends to participate. By accounting for features of the social networks and recruitment patterns generated by this seemingly biased and ad hoc sampling strategy, Salganik and Heckathorn (2004) have shown that asymptotically unbiased estimates of population characteristics can be quickly and efficiently generated with fairly modest assumptions about the extent to which all members of the population are linked to each other through friendship (or acquaintance) networks.

The promise of respondent-driven sampling for understanding features of the undocumented migrant population has recently been suggested by a series of studies using the technique to evaluate labor law violations among populations of low-wage workers, many of whom may be undocumented (see, e.g., Bernhardt, Polson, and DeFilippis, 2010).

There is still the problem of encouraging undocumented migrants to provide accurate information about their immigration experience. However, because the chief interest is in population estimates rather than information about individual experiences, there are good candidate procedures for minimizing respondents’ anxieties about providing correct information. For instance, GAO (2006) has endorsed the “grouped answers” approach to surveying foreign-born respondents. With this approach, individual respondents never have to acknowledge any illicit activities because the illicit activities they are engaged in are grouped with licit activities, and they choose the group of activities that contains the correct answer for themselves. For instance, half the respondents might be asked to say if they are either



- A. a U.S. citizen or
- B. a legal permanent resident, currently undocumented, or a refugee or asylee.

The other respondents are asked to say whether they are

- A. a legal permanent resident, refugee, or asylee or
- B. a U.S. citizen or currently undocumented.

In this example, the percentage of undocumented migrants can be calculated by subtracting the percentage of foreign-born respondents who answer B in one question set, from the percentage who answer A in the other set. GAO (2006) reports tests showing that respondents are generally comfortable answering grouped questions that combine licit with illicit behaviors in such a way that respondents never reveal their individual status.

## Synthetic and Proxy Measures

There are some quantities important to understanding government performance that simply cannot be measured directly and which require instead some more-complex modeling to generate even plausibly correct estimates. For instance, for many years the Office of National Drug Control Policy (ONDCP) published estimates of the amount of money U.S. residents spend on illicit drugs (ONDCP, 2001). Generating these estimates entailed complex modeling and assumptions integrating disparate sources of information on, for instance, the prevalence of heavy drug use among arrestees, the proportion of the population that has been arrested, the amount of drugs typically consumed by heavy and light users, the amount of time heavy users spend incarcerated so not at risk of drug use, the typical street price of drugs, etc. By combining these multiple parameter estimates in a theoretically logical way, ONDCP produced reasonably credible price series showing how drug spending was rising or falling over time.

Similar approaches to generating synthetic estimates may be possible for establishing the flow of undocumented immigrants between ports of entry. For instance, estimates of gross annual illegal migration, such as those produced by Passel, Van Hook, and Bean (2004), could be reduced by estimates of the number of such migrants arriving through ports of entry or overstaying their visas, values that would have to be separately generated from other data sources and models. Similarly, improving the capture-recapture approach discussed above would likely require developing a model of migrant flows in and out of the country, with multiple data sources contributing to estimates of the portion of apprehended migrants who are deterred from further attempts, the proportion who successfully cross between separate apprehension events, the subgroup makeup of migrants over time, and other important characteristics. Potentially good sources of data exist that have not yet been mined for these purposes, including the EMIF (described above) and the Mexican National Survey of Employment and Occupation (ENOE), a representative household survey that explicitly identifies household members who have departed or arrived in the last 90 days.

Bennett, Stewart, and Malmin (2001) offered one example of a synthetic flow model. They used sequential-year Mexican census results to identify annual changes to population demographic characteristics that could not be explained by births or deaths and identified these as emigration risk factors. Separately, they constructed a regression model linking changes in the

size of the at-risk population to economic and other factors likely to induce the at-risk population to emigrate. The results of these analyses were annual series illustrating the size of the at-risk population in Mexico and model-based estimates for the proportion of that population that would seek entry to the United States annually. This demand model could in principle be combined with estimates about the number of Mexican immigrants passing through ports of entry to generate a reasonable estimate of flow between ports of entry.

Similarly, Wein, Liu, and Motskin (2009) develop a mathematical model of illegal immigration with model components to produce estimates of the likelihood of attempting a crossing, the probability of apprehension, the probability of being released into the United States despite being apprehended, and labor market demand for illegal immigrants in the United States. Although this model has not been developed to provide timely estimates of the flow of migrants between ports of entry, it represents a promising candidate for such development.

Proxy measures too rely on an implicit or explicit model relating the phenomenon of interest to the proxy. For instance, the cost of coyote services bears some relationship to the effectiveness of border control efforts and the flow of migrants between ports of entry. Cornelius (2001), for instance, relates short-term price changes in coyote services to concurrent enhancements in border protection efforts. But as an ongoing border control measure, the relationship between prices and border control may be insufficiently straightforward, because prices are also sensitive to a range of other factors that must be accounted for, such as

- the risks and other costs of providing coyote services, which may fluctuate with changes in Mexican law and enforcement efforts, U.S. laws and enforcement, and hazards associated with competition between rival coyote service providers
- the costs of attempting a crossing without the aide of a coyote, which may fluctuate with changes in border control policies and procedures, the weather, and other factors
- the migration alternatives other than crossing between ports of entry that are available to would-be crossers
- the size and structure of the coyote service industry, since, for instance, greater competition for clientele could reduce prices, and changes to the structure of the industry, such as the formation of more professional and consolidated service provider organizations, could similarly affect pricing
- migrants' ability to afford the services, because as more migrants can afford to pay coyotes, demand increases, and prices are likely to rise
- the specific services the coyote is contracted to provide, since, for example, migrants typically pay more to be smuggled to the interior in a tractor-trailer than they pay a guide to hike around checkpoints in the hot Texas brush, and more to be smuggled through a port of entry with false documents and/or with the cooperation of a corrupt inspector than to be taken across the Rio Grande in an inner tube.

These multiple effects on the price of coyote services make clear that price series alone cannot be used to estimate either flow or border control performance. However, it may be possible to develop economic models that account for the multiple influences on coyote pricing in such a way that flow could be more reliably estimated.

## Discussion

---

The objectives of this paper were to review a selection of existing and candidate methods for estimating the flow of people crossing into the United States between ports of entry and to offer DHS suggestions on how to proceed in developing reliable and valid measures of such flow for use in measuring border control performance. Flow of illegal crossers is a key measure of the effectiveness of U.S. border control efforts, but we find that current approaches to estimating flow are not sufficiently precise to serve as useful performance measures. Nevertheless, as this paper highlights, there are several promising approaches that merit development. These include

- further development of the capture-recapture methods to include reliable estimates of attrition in the pool of migrants attempting to cross after an apprehension and of the number and length of successful crossings migrants are making between apprehensions
- estimating total crossings from crossings in a stratified sample of border segments
- respondent-based sampling surveys of migrant communities in the United States
- synthetic modeling either of the stock and flow of the migrant community or of the black market economy for coyote services.

Moreover, each of these approaches might beneficially be used in conjunction with the others. Capture-recapture methods could be used to inform synthetic model parameter estimates for the probability of apprehension or discouragement. Respondent-driven sampling results could be used to examine whether there are subgroups of migrants with higher or lower risks of apprehension, information that could inform both improved capture-recapture and synthetic models. And stratified sampling could be used to examine how flow is shifted from one sector to another when border control efforts are enhanced in the first, information vital to understanding whether cases that appear deterred or “discouraged” are, in fact, merely displaced.

If CBP systematically varied the probability of patrols being present in a region of the border as part of a stratified sampling plan, changes in recapture rates in those regions could be used to estimate the deterrence effects of increased patrols. For instance, suppose that after a 10 percent increase in the probability of patrols at a given segment of border, there is only a 5 percent increase in apprehensions of crossers who were recently repelled on a first attempt to enter the country. If we assume the number of apprehensions of crossers making a second attempt is simply the product of the probability of a patrol being present, a fixed probability of apprehension given the presence of a patrol, and the number of crossers making a second attempt, then the difference between the observed number of second-time apprehensions and



the expected number of such apprehensions (accounting for any changes in the number of migrants apprehended once) could be attributable to a deterrent effect.

Measurement of illegal cross-border flows is a rich challenge with many opportunities for innovation. As such, it is a problem that could benefit from the focused attention of a broad scientific community, such as might be cultivated were DHS to offer grant funding awarded competitively to support new thinking and analysis on this topic.

Moreover, because of the inherent complexities of producing reliable estimates of illegal cross-border movements and uncertainties about whether any particular approach is likely to succeed, DHS should consider developing several alternative approaches to flow measurement concurrently. By producing independent estimates from distinct methods and data sources, DHS will have far more confidence in any resulting comparability in the estimates, and differences between them will highlight uncertainties that managers must consider as they plan for and resource future border control efforts.

## References

---

- Bennett, M. J., Stewart, J. L., Malmin, O. K. (2001). *An Assessment of the Southwest Border Enforcement*. Alexandria, Va.: CNA Corporation.
- Bernhardt, A., Polson, D., DeFilippis, J. (2010). *Working Without Laws: A Survey of Employment and Labor Law Violations in New York City*, New York: National Employment Law Project. As of September 24, 2010: <http://www.nelp.org/WorkingWithoutLawsNYC>
- CBP—see U.S. Customs and Border Protection.
- Chang, J. C., Kohout, R., Blacksten, H. R., Benard, S. (2006). *CBP Apprehensions at the Border*, Arlington, Va.: Homeland Security Institute, not available to the general public.
- Cornelius, W. A. (2001). “Death at the Border: Efficacy and Unintended Consequences of U.S. Immigration Control Policy,” *Population and Development Review*, Vol. 27, pp. 661–685.
- DHS—see U.S. Department of Homeland Security.
- Donato, K. M., Durand, J., Massey, D. S. (1992). “Stemming the Tide? Assessing the Deterrent Effects of the Immigration Reform and Control Act,” *Demography*, Vol. 29, No. 2, pp. 139–157.
- Eschbach, K., Hagan, J. M., Rodríguez, N. P., Hernández-León, R., Bailey, S. (1999). “Death at the Border,” *International Migration Review*, Vol. 33, No. 2, pp. 430–440.
- Espenshade, T. J. (1995). “Using INS Border Apprehension Data to Measure the Flow of Undocumented Migrants Crossing the U.S.–Mexico Frontier,” *International Migration Review*, Vol. 29, pp. 545–565.
- GAO—see U.S. Government Accountability Office.
- International Trade Administration: Office of Travel and Tourism Industries (undated). “2009 Monthly Tourism Statistics, Table C—Section 1: Total Visitation, Canada, Mexico, Total Overseas, Western Europe Non-Resident Visitation to the U.S. by World Region/Country of Residence 2009.” As of September 20, 2010: <http://tinet.ita.doc.gov/view/m-2009-I-001/table1.html>
- ITA—see International Trade Administration.
- Office of National Drug Control Policy (December 2001). *What America’s Users Spend on Illegal Drugs*, Washington, D.C.
- ONDCP—see Office of National Drug Control Policy.
- Passel, J. S. (June 14, 2005). “Unauthorized Migrants: Numbers and Characteristics; Background Briefing Prepared for Task Force on Immigration and America’s Future,” Washington, D.C.: Pew Hispanic Center.
- Passel, J. S., Cohn, D. (September 1, 2010). *U.S. Unauthorized Immigration Flows Are Down Sharply Since Mid-Decade*, Washington, D.C.: Pew Hispanic Center.
- Passel, J. S., Van Hook, J., Bean, F. D. (June 1, 2004). *Estimates of the Legal and Unauthorized Foreign-Born Population for the United States and Selected States, Based on Census 2000*, Washington, D.C.: Urban Institute, Report to the Census Bureau. As of January 5, 2011: [http://www.copafs.org/seminars/methodology\\_and\\_data\\_quality.aspx?](http://www.copafs.org/seminars/methodology_and_data_quality.aspx?)
- Predd, J. B., Willis, H. H. (2010). *The Role of Pattern and Trend Analysis in Border Control*, unpublished RAND Corporation research.

Prochaska, J. O., DiClemente, C. C., Norcross, J. C. (September 1992). "In Search of How People Change; Applications to Addictive Behaviors," *American Psychologist*, Vol. 47, No. 9, pp. 1102–1114.

Rendall, M. S., Aguila, E., Basurto-Dávila, R., Handcock, M. S. (May 7, 2009). *Migration Between Mexico and the U.S. Estimated from a Border Survey*, prepared for presentation at the 2009 Annual Meeting of the Population Association of America, Detroit, and the 2009 Metropolis British Columbia and Center for Research and Analysis of Migration Workshop on the Economics of Immigration. As of September 18, 2010: <http://mbc.metropolis.net/assets/uploads/files/Rendallnew.pdf>

Salganik, M. J., Heckathorn, D. D. (2004). "Sampling and Estimation in Hidden Populations Using Respondent-Driven Sampling," *Sociological Methodology*, Vol. 34, pp. 193–239.

Thompson, W. L. (2004). *Sampling Rare or Elusive Species: Concepts, Designs, and Techniques for Estimating Population Parameters*, Washington, D.C.: Island Press.

U.S. Customs and Border Protection (November 24, 2009). "Securing America's Borders: CBP Fiscal Year 2009 in Review Fact Sheet," Washington, D.C. As of August 7, 2010: [http://www.cbp.gov/xp/cgov/newsroom/news\\_releases/archives/2009\\_news\\_releases/nov\\_09/11242009\\_5.xml](http://www.cbp.gov/xp/cgov/newsroom/news_releases/archives/2009_news_releases/nov_09/11242009_5.xml)

——— (July 2010). "Snapshot: A Summary of CPB Facts and Figures," Washington, D.C. As of September 20, 2010: <http://www.cbp.gov/linkhandler/cgov/about/accomplish/snapshot.ctt/snapshot.pdf>

U.S. Department of Homeland Security (February 2010). *Quadrennial Homeland Security Review Report: A Strategic Framework for a Secure Homeland*, Washington, D.C. As of September 20, 2010: [http://www.dhs.gov/xlibrary/assets/qhsr\\_report.pdf](http://www.dhs.gov/xlibrary/assets/qhsr_report.pdf)

U.S. Government Accountability Office (July 2005). *Border Patrol: Available Data on Interior Checkpoints Suggests Differences in Sector Performance*, Washington, D.C., GAO-05-435.

——— (September 2006). *Estimating the Undocumented Population: A "Grouped Answers" Approach to Surveying Foreign Born Respondents*, Washington, D.C., GAO-06-775.

——— (November 2007). *Border Security: Despite Progress, Weaknesses in Traveler Inspections Exist at Our Nation's Ports of Entry*, Washington, D.C., GAO-08-219.

——— (May 2008). *Border Security: Summary of Covert Tests and Security Assessments for the Senate Committee on Finance, 2003–2007*, Washington, D.C., GAO-08-757.

——— (August 2009). *Border Patrol: Checkpoints Contribute to Border Patrol's Mission, But More Consistent Data Collection and Performance Measurement Could Improve Effectiveness*, Washington, D.C., GAO-09-824.

Wein, L. M., Liu, Y., Motskin, A. (May 2009). "Analyzing the Homeland Security of the U.S.-Mexican Border," *Risk Analysis*, Vol. 29, No. 5, pp. 699–713.

Willis, H. H., Predd, J. B., Davis, P. K., Brown, W. (2010). *Measuring the Effectiveness of Border Security Between Ports-of-Entry*, Santa Monica, Calif.: RAND Corporation, TR-837-DHS. As of January 6, 2011: [http://www.rand.org/pubs/technical\\_reports/TR837.html](http://www.rand.org/pubs/technical_reports/TR837.html)